

Conference Abstract

Importance of Language Localization: Attempts for Collecting and Publishing Specimen Data in Japan

Aino T. Ota[‡], Minako Abe Ito[‡], Showtaro Kakizoe[‡], Utsugi Jinbo^{‡,§}

[‡] Center for Collections, National Museum of Nature and Science, Tsukuba, Japan

[§] Department of Zoology, National Museum of Nature and Science, Tsukuba, Japan

Corresponding author: Aino T. Ota (mt.indigo@gmail.com)

Received: 15 Nov 2024 | Published: 18 Nov 2024

Citation: Ota AT, Ito M, Kakizoe S, Jinbo U (2024) Importance of Language Localization: Attempts for Collecting and Publishing Specimen Data in Japan. Biodiversity Information Science and Standards 8: e141888. <https://doi.org/10.3897/biss.8.141888>

Abstract

Language localization is a process of adapting the needs of a certain linguistic community, including its cultural contexts, to adjust the nuance and wording of the original language. It is not just a translation from English, the de facto language of publication in the scientific community. This is particularly important in a linguistically diverse region such as Asia, including Japan, and thus language localization is crucial for developing an environment that enables the collection and utilization of biodiversity-related data in their first language. The Japan Biodiversity Information Initiative ([JBIF](#)) is an organization aiming to carry out activities of the Global Biodiversity Information Facility ([GBIF](#)) in Japan. In this presentation, we will introduce our experiences of language localization in JBIF, which includes improving access, adapting Japanese names (a kind of vernacular name), and images.

JBIF has been working to improve the Japanese community's access to GBIF data, translating the GBIF website, and a variety of other activities. These works are related to one of our missions, "utilization of biodiversity data." Furthermore, Science Museum Net ([S-Net](#)), a portal site operated by the National Museum of Nature and Science ([NMNS](#)), has been collecting natural history specimen data from research institutions and

museums, including NMNS itself; has disseminated data through GBIF; and has provided data searchability for domestic users.

Language localization has contributed to internationally disseminating biodiversity information by improving access to global platforms. S-Net has adopted a data format that includes items written in Japanese and corresponding occurrence terms in the Darwin Core Standard (Darwin Core Task Group 2009), as well as several unique fields requested by domestic users. S-Net also designed the data entry flow, which allows the entire process—from submission to domestic and/or international publication—to be completed in Japanese. With this data entry process, S-Net has collected more than 7.3 million data records, the largest number in Asia, from more than 100 museums and institutions over the past 20 years. Along with government support, language localization inevitably helps reduce barriers for many participants, including small museums with limited capacity to submit specimen data.

In this context, language localization is particularly crucial for the treatment of species names. In the Japanese community, many organisms have “Japanese names” that corresponds to each taxon. As Japanese names are familiar to both the scientific community and citizens in Japan, Japanese names are more widely used than scientific names. Therefore, scientific and citizen communities must be able to access biodiversity information using Japanese names. JBIF has provided a service to the GBIF database by converting the Japanese name to a scientific name using the species name dictionary. This required the development of a species list that includes Japanese names in addition to scientific names and serves as the backbone of the database. This localization could also apply to several Asian countries where there is a vernacular name system similar to that of Japan.

Language localization also plays an important role in developing databases that include accompanying images with biodiversity information, a growing trend in Japan. With the amendment of the Museum Act in 2023, museums in Japan are now required to digitize and publish specimens and materials in their possession. By publishing and sharing specimen images, various uses beyond science are expected. Some museums, including NMNS, already have databases that can publish accompanying images with text data, while other museums, especially small museums, do not. S-Net is expected to add a function to publish images. S-Net is also connected to an integrated platform for searching aggregated metadata and images in various fields in Japan, such as data from museums, galleries, and libraries.

Images should also contribute to improving the accuracy of biodiversity-related data by complementing language localization. During digitization, typing errors and misinterpretations may occur, and raw data such as handwritten labels may be lost. Publication of such accompanying information in images, other than the specimen itself, should allow data users to validate accuracy and identify potential errors to be corrected.

S-Net should expand its role from improving image datasets by developing an image release system. To achieve this, it is essential to cultivate a culture of Findability,

Accessibility, Interoperability and Reuse (FAIR) and open data, and to establish Creative Commons licensing and data handling protocols among stakeholders in Japan.

We believe that our work serves as a valuable precedent for collection and dissemination of biodiversity information from Asia, and for fostering collaboration among research communities, citizens, and governments at appropriate scales in Asian countries and the rest of the world.

Keywords

JBIF, GBIF, Asia, Japanese name, database, S-Net

Presenting author

Aino T. Ota

Presented at

SPNHC-TDWG 2024

Funding program

This work was supported by the National BioResource Project (NBRP) by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Darwin Core Task Group (2009) Darwin Core. Biodiversity Information Standards (TDWG). URL: <http://www.tdwg.org/standards/450>